AFRL-SR-AR-TR-10-0025

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
24-02-2010	Final Performance Report	March 01, 2007 -- November 30, 2009

4. TITLE AND SUBTITLE
New Graph Models and Algorithms for Detecting Salient Structures from Cluttered Images

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-07-1-0250

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Wang, Song

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of South Carolina
901 Sumter St, STE 511
Columbia, SC 29208

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Office of Scientific Research
875 North Randolph Street
Suite 325, Rm 3112
Arlington, VA 22203

NE

10. SPONSOR/MONITOR'S ACRONYM(S)
AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution A: Approved for Public Release

20100504283

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This research was focused on developing new efficient algorithms to automatically detect structures of interest from cluttered images. The major accomplishments include: 1) a unified framework for edge grouping that can detect both open and closed boundaries from cluttered images, 2) a new graph model and algorithm to detect perceptually salient structures that show a certain level of boundary symmetry from cluttered images, 3) a new MCMC-based partial shape matching algorithm that can match two 2D shape contours with nonrigid shape deformation and multiple partial occlusions, 4) a new free-shape subwindow search algorithm for object localization that outperforms the state-of-the-art rectangle subwindow search algorithms, 5) two new perceptually motivated strategies for shape classification and recognition that leads to the new state-of-the-art recognition rate on the widely used MPEG7 shape data set, and 6) a new benchmark for more objective shape-correspondence performance evaluation along with a new shape-correspondence algorithm for statistical shape analysis that pre-organizes the given population of shape instances to achieve high correspondence accuracy using less CPU time.

15. SUBJECT TERMS
Image processing, image segmentation, edge grouping, shape analysis, shape matching, object localization

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	20	

AFOSR Final Performance Report

Project Title:
New Graph Models and Algorithms for Detecting Salient Structures from Cluttered Images

Award Number:
FA9550-07-1-0250

Dates Covered:
March 01, 2007 – November 30, 2009

Program Manager:
Dr. Jon A. Sjogren
Air Force Office of Scientific Research (AFOSR/NE)
Physics and Electronics
875 North Randolph Street
Suite 325 Room 3112
Arlington, Virginia 22203
Tel (703) 696 6564
Fax (703) 696 8450
Email: jon.sjogren@afosr.af.mil

Principal Investigator:
Song Wang
Department of Computer Science and Engineering
University of South Carolina
Columbia, SC 29208
Tel: (803) 777 2487
Fax: (803) 777 3767
Web: http://www.cse.sc.edu/~songwang
Email: songwang@cec.sc.edu

February 24, 2010

Summary

In this project, we conducted research on developing new models and algorithms to address the fundamental edge grouping problem in computer vision and image processing. Based on edge grouping results, we further developed new partial shape matching, object localization, shape-based classification, and shape correspondence algorithms to detect structures or objects of interest from cluttered images. The major accomplished work includes:

1. Development of a unified framework for edge grouping that can detect both open and closed boundaries from a cluttered image. A closed boundary corresponds to the case in which the desirable object is completely located within the image perimeter, while an open boundary corresponds to the case in which the desirable object is partially cropped by the image perimeter. In this framework, a set of edge and region features are first detected from the image. These features are then integrated into a unified grouping cost (a measure negatively related to the structural saliency) that takes a ratio form: the numerator describes the edge features and the denominator describes the region features. We found that the globally optimal boundary that minimizes this unified grouping cost can be found in polynomial time by using graph models and algorithms.

2. Development of graph models and algorithms to detect boundaries that show certain levels of symmetry, an important geometric property of many structures of interest. We addressed this problem by encoding boundary symmetry into edge grouping. More specifically, we constructed a new grouping token by pairing the detected edges into some symmetric trapezoids and some gap-filling quadrilaterals. Based on these, we defined a grouping cost that incorporates a term for boundary symmetry and constructed a graph model in which a symmetric boundary can always be modeled by a path. Finally, we adapted our graph models and algorithms for finding the optimal path corresponding to the desirable symmetric boundary.

3. Development of a new partial shape matching algorithm to match two 2D contours with mild nonrigid shape deformation and multiple partial occlusions. This algorithm identifies and matches a subset of fragments of the two contours and finds the one-to-one dense point correspondence between them. More specifically, we used the MCMC (Markov chain Monte Carlo) algorithm to search for the matched subset of fragments. This partial shape matching algorithm can be used for matching detected boundaries (resulting from edge grouping) against a set of per-stored template object boundaries for object detection and segmentation.

4. Development of a free-shape subwindow search algorithm for object localization. We adapted the graph models and algorithms developed for edge grouping for localizing the objects of interest by finding a tighter free-shape covering subwindow. The state-of-the-art bag of visual words technique is used to detect, describe, and quantify the features, where the desirable object features and the background features are distinguished by using supervised SVM (support vector machine) learning. We tested the developed algorithm on the widely-used PASCAL VOC2006 and PASCAL VOC2007 databases, where each category of objects bears very large within-category variations. We found that the performance of the developed algorithm is better than the current state-of-the-art efficient subwindow search algorithms.

5. Development of two perceptually motivated strategies for shape classification and recognition. The first strategy handles shapes that can be decomposed into a base structure and a set of inward or outward pointing strand structures, where a strand structure represents a very thin, elongated shape part attached to the base structure. We decomposed such shapes and

computed their shape similarities by measuring the similarity of their base structures and strand structures separately. The second strategy handles shapes that exhibit good bilateral symmetry. We developed an algorithm to identify such symmetric shapes and unify their aspect ratio in terms of their symmetry axis before measuring the shape similarity. We found that these two strategies can be integrated into available shape matching methods to achieve the new state-of-the-art classification performance on the widely-used MPEG7 shape dataset.

6. Development of a new benchmark for shape-correspondence performance evaluation. Different from previous shape-correspondence evaluation methods, the proposed benchmark first generates a large set of synthetic shape instances by randomly sampling a given statistical shape model that defines a ground-truth shape space. The proposed benchmark allows for a more objective evaluation of shape correspondence than previous methods. We also developed a new shape correspondence algorithm that pre-organizes the population of shape instances in a tree, where each node represents a shape instance and each edge connects two very similar shape instances. We then only correspond shape-instance pairs that are connected by an edge. Testing on the benchmark shows that the new algorithm achieves high correspondence accuracy and low algorithm complexity simultaneously.

1 Edge Grouping for Open and Closed boundaries

For edge grouping, we first detect a set of edges, as shown in Fig. 1(b), from an input image $I(x, y)$, as shown in Fig. 1(a). We refer to these edges as *detected segments*. Second, we construct an additional set of straight line segments to connect every pair of detected segments. We refer to these new straight line segments as *gap-filling segments*. A *closed* boundary is then defined as a cycle of alternating detected and gap-filling segments, as shown in Fig. 1(d). To unify both open and closed boundary detection in the grouping, we divide the image perimeter into a set of detected segments, as shown in Fig. 1(e). If the resulting optimal closed boundary contains one or more segments constructed from the image perimeter, it is actually an open boundary, as shown in Fig. 1(f), where the resulting boundary (red thick segments) contains part of the perimeter and in fact is an open boundary.

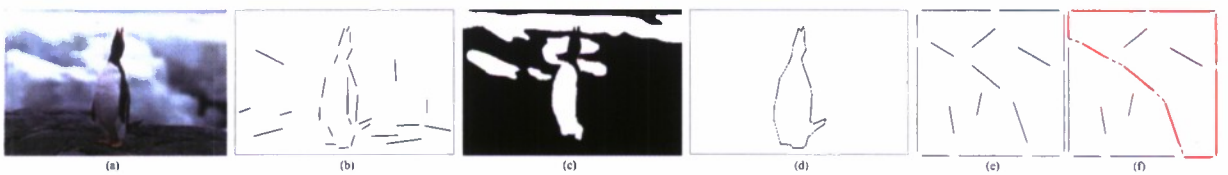

Figure 1: An illustration of edge grouping for unified closed and open boundary detection. (a) Input image, (b) detected segments, (c) binary feature map, (d) the detected *closed* boundary that traverses detected (solid) and gap-filling (dashed) segments alternately, (e) dividing the image perimeter into a set of detected segments, and (f) when the detected closed boundary (red thick segments) contains part of the image perimeter, it represents an open boundary cropped by image perimeter.

Region information can be integrated into edge grouping by constructing a binary feature map as shown in Fig. 1(c). A binary feature map $M(x, y)$ is of the same size as the input image $I(x, y)$ and reflects whether pixel (x, y) has a desired property or not. It can be constructed from the input image $I(x, y)$ using an image-analysis method and/or any available *a priori* knowledge of the appearance of the desirable salient structures. We set $M(x, y) = \alpha$ (white) to indicate that pixel (x, y) belongs to the desired structure and $M(x, y) = \beta$ (black) otherwise. Note that feature map also contains

noise and errors. We set $\alpha > 0$ and $\beta < 0$ such that $\sum_{(x,y)} M(x,y) = 0$. Without loss of generality, we set $\alpha = 1$ and

$$\beta = -\frac{\sum_{(x,y):M(x,y)>0} 1}{\sum_{(x,y):M(x,y)<0} 1}. \tag{1}$$

We defined a unified grouping cost for a candidate (open or closed) boundary $\mathcal{B}$ as

$$\phi(\mathcal{B}) = \frac{|\mathcal{B}_G|}{\left| \iint_{R(\mathcal{B})} M(x,y)dxdy \right|}, \tag{2}$$

where $|\mathcal{B}_G|$ is the total length of all the gap-filling segments along the boundary $\mathcal{B}$. This accounts for the Gestalt law of proximity, where a smaller total gap length $|\mathcal{B}_G|$ represents better proximity. $R(\mathcal{B})$ is the region enclosed by the boundary $\mathcal{B}$ and $\iint_{R(\mathcal{B})} M(x,y)dxdy$ is the sum of the feature values of the pixels, taken from the binary feature map M, inside the region enclosed by $\mathcal{B}$. We found that that the ratio-contour algorithm [42] can be used to find the global optima of this grouping cost.

Sample experimental results are shown in Fig. 2. This algorithm can be applied to the segmented regions recursively to obtain a hierarchical image segmentation, as shown in Fig. 3. The hierarchical image segmentatiton performance on a selected 100 natural images from the Berkeley dataset is shown in Table 1, with comparisons to several other state-of-the-art image segmentation algorithms. In this performance evaluation, we use the boundary consistency measure [27] in the Berkeley Benchmark, which provides precision, recall, and an integrated "F-measure" of the detected boundaries against the manual segmentation.

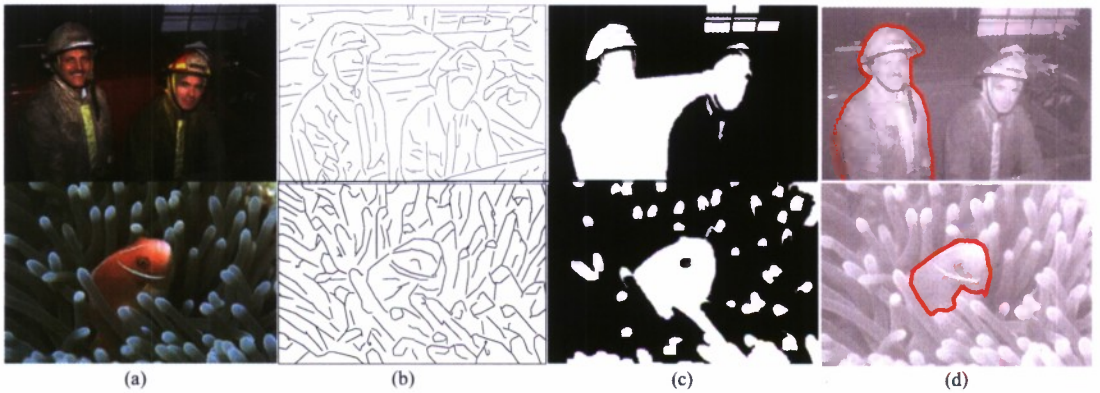

Figure 2: Sample experimental results of the edge grouping for both open and closed boundaries. (a) Input images, (b) detected segments, (c) constructed binary feature map, and (d) detected boundaries.

2 Edge Grouping for Symmetric boundaries

Structures of interest encountered in many real images show a certain level of (bilateral) symmetry over a particular axis. The straighter the symmetry axis, the higher the symmetry of the underlying structural boundary. We developed new graph models and algorithms to address symmetric edge grouping in a globally optimal fashion. Based on the edge grouping framework described in Section 1,

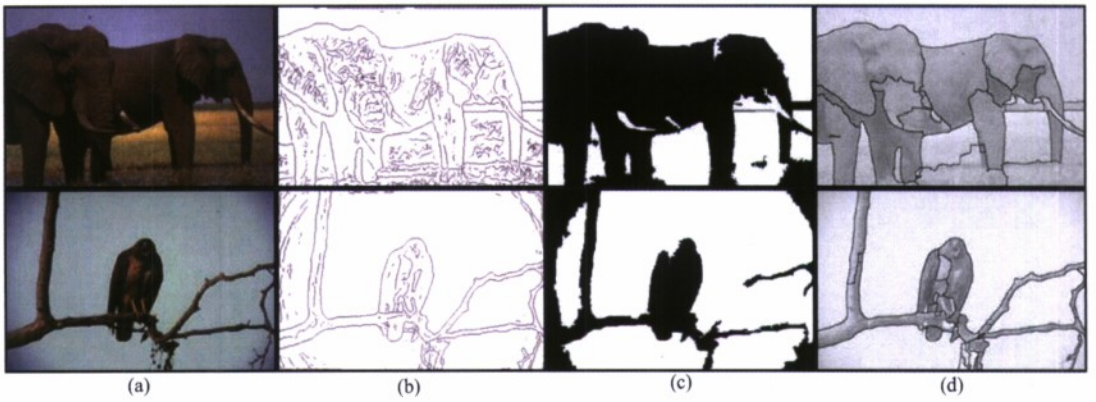

Figure 3: Sample experimental results of the iterative edge grouping for hierarchical image segmentation. (a) Input image, (b) detected segments, (c) binary feature map constructed for the first iteration of edge grouping, and (d) resulting image segmentation.

Method	Recall	Precision	F-measure
Berkeley Edge Detector [26]	**0.7058**	0.6857	**0.6956**
Proposed Method	0.6597	**0.6973**	0.6780
Ultrametric Contour Maps [4]	0.6860	0.6576	0.6715
BGCGTG [27]	0.6934	0.6078	0.6478
Statistical Region Merging (SRM) (Q = 128) [32]	0.6989	0.5241	0.5990
Linear Multiscale Normalized Cut [12]	0.5940	0.5787	0.5862

Table 1: Image segmentation performance on the Berkeley Benchmark [28, 27], according to a boundary-consistency measure. Note that while Berkeley Edge Detector [26] shows a higher F-measure value, it only detects incomplete boundaries and cannot accomplish a region-based segmentation.

we constructed a new grouping token by pairing the detected segments into some symmetric trapezoids, as shown in Figs. 4(a) or (b), where three trapezoids $\mathcal{T}_1 = \{P_1P_2P_{11}P_{12}\}$, $\mathcal{T}_2 = \{P_3P_4P_9P_{10}\}$, and $\mathcal{T}_3 = \{P_5P_6P_7P_8\}$ are constructed from detected-segment pairs P_1P_2 & $P_{11}P_{12}$, P_3P_4 & P_9P_{10}, and P_5P_6 & P_7P_8 respectively. Q_1Q_2, Q_3Q_4 and Q_5Q_6 are the symmetry axes of these three trapezoids. To group these trapezoids into a closed boundary, we constructed some quadrilaterals to fill the gap between the trapezoids. For the examples shown in Figs. 4(a) and (b), two gap-filling quadrilaterals $\mathcal{G}_1 = \{P_2P_3P_{10}P_{11}\}$ and $\mathcal{G}_2 = \{P_4P_5P_8P_9\}$ can connect the three trapezoids $\mathcal{T}_1$, $\mathcal{T}_2$ and $\mathcal{T}_3$ into a closed boundary $\mathcal{B} = P_1P_2 \ldots P_{12}P_1$ with a polyline axis $\texttt{axis}(B) = Q_1Q_2 \ldots Q_6$. We defined a grouping cost for such a boundary $\mathcal{B}$ as

$$\phi(\mathcal{B}) = \frac{|\mathcal{B}_G| + \lambda \cdot \rho(\texttt{axis}(\mathcal{B}))}{\texttt{area}(\mathcal{B})}, \tag{3}$$

where $\rho(\texttt{axis}(\mathcal{B}))$ is a measure of the straightness of the polyline axis of the boundary $\mathcal{B}$.

We then constructed a graph where each vertex represents a trapezoid axis endpoints, i.e., $Q_i, i = 1, 2, \ldots, 6$ in Fig. 4 and each solid (or dashed) edge represents a trapezoid (or quadrilateral) axis. By embedding the grouping cost (3) to edge weights, we reduced the edge-grouping problem to a problem of finding an alternate path with a minimum ratio-form cost in this graph. We found that the ratio-contour algorithm can be adapted to find such a globally optimal path in polynomial time. Sample results are shown in Fig. 5.

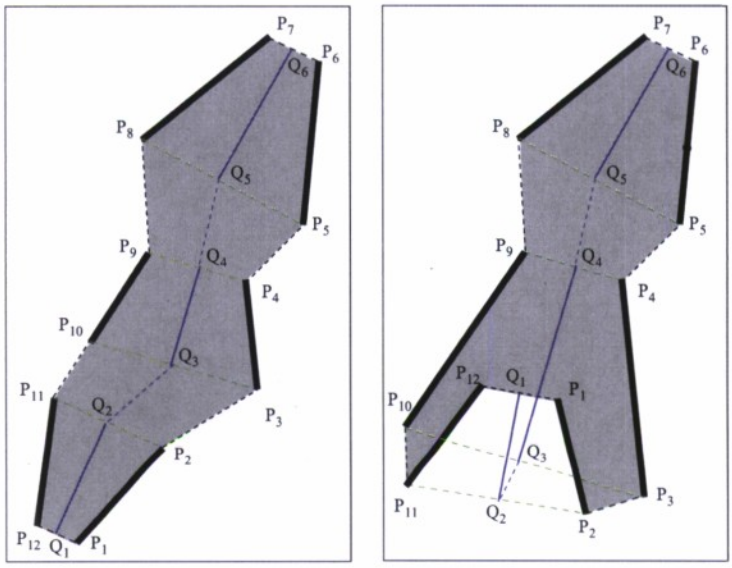

Figure 4: An illustration of grouping detected trapezoids into a closed boundary.

3 Partial Shape Matching

We developed a new algorithm for partial shape matching that can better handle nonrigid shape deformation and allow the matching of multiple disjoint contour fragments. We represent each contour by a sequence of landmark points and the partial shape matching is reduced to a problem of selecting subsequences of these landmark points and matching them. We used the MCMC (Markov Chain Monte Carlo) technique [18] to find the globally optimal matching.

Using Bayesian inference, we set the goal to find a partial shape matching with the maximal posterior probability, which is the product of *likelihood* and *prior*. The likelihood describes the matching cost between the selected landmarks on the two contours and the prior specifies certain general preference on the landmark selection on each contour. To account for the nonrigid shape deformation between them, we defined the likelihood using the Procrustes distance [15] between

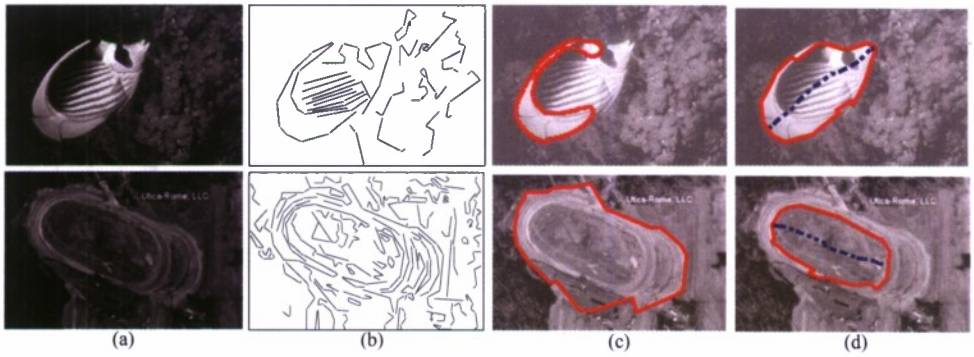

Figure 5: Sample edge grouping results considering symmetry information. (a) Input image, (b) detected segments, (c) edge grouping result without considering symmetry, and (d) edge grouping result by considering symmetry.

the selected subsequences of landmark points on these two contours: the smaller the Procrustes distance, the larger the likelihood. Procrustes distance is invariant to rotation, translation and scaling transforms [7].

We considered the following gap penalty prior density in the form of $e^{(-a \cdot GN - b \cdot GL)}$, where $a > 0$ and $b > 0$ are the penalty parameters, GN is the number of gaps, and GL is the total gap length. A 'gap' is defined as either a single unmatched landmark point or a set of consecutive unmatched landmark points. To avoid counting the gaps that are introduced by denser landmark sampling on a contour, we sample sparser landmark point on one contour than on the other and then measure GN and GL on the former contour. This prior provides two desirable properties. First, the larger the number of landmark points selected for matching, the better. The second desirable property is that we prefer un-selected landmark points to occur in sequence rather than being spread over the contour. Without this property, the algorithm may favor many short and disjoint matching contour fragments.

With this prior and the likelihood, we estimated the posterior for any partial shape matching results using an MCMC inference, such as the Metropolis-Hastings algorithm [30], to search for the optimal matching. The effectiveness of an MCMC scheme is highly dependent on the choice of proposal distribution. We used two simple proposals in our algorithm: (i) the *match-unmatch* proposal, where a randomly selected point is removed from the matched subsequence if it is currently in the matched subsequences, and vice versa, with a given probability; and (ii) the *match-match* proposal, where a randomly selected point, which is currently in the matched subsequence (matched to a landmark point on the other contour), is set to match a new randomly selected point on the other contour without breaking contour topology, i.e., when connecting the identified landmarks on each contour with the specified order, no self-intersection will be produced.

Sample results using this partial shape matching algorithm are shown in Fig. 6. We also quantitatively compare the performances of this MCMC-based algorithm with the performance of the Smith-Waterman algorithm used in [10] on 40 sets of synthetic contour pairs. Each contour set consists of 40 shape-contour pairs that are constructed from the well-known MPEG7 shape dataset [22] by introducing various nonrigid shape deformations and partial shape occlusions. The comparison results are shown in Fig. 7, where the matching score is based on the coincidence between the obtained and the ground-truth partial shape matchings. This matching score penalizes both false positive and false negative matched fragments.

4 Free-shape Subwindow Search for Object Localization

Localizing objects with large within-category variation requires effective methods to (a) identify good features to distinguish the objects of interest from the cluttered background, and (b) search for the regions that show strong features identifying objects of interest.

The bag of visual words technique [41, 19, 23] is the current state of the art technique for feature detection. In this technique, a large set of image features are detected and quantified into a small set of visual words. A classifier is trained on training images (with labeled foreground object and background) to associate a feature score with each visual word: if a visual word bears a positive score, it is more like a feature of the desirable object of interest, and if a visual word bears a negative score, it is more like a feature of the background. Sliding window [9, 17] is a widely-used technique for feature based object localization: for every possible subwindow in an image, the feature scores covered by the window are checked and the one with the maximum total feature scores is selected as the optimal subwindow as thus the location of the object. Recently, more efficient branch and bound algorithms [21, 3] have been developed to speed up the subwindow search without exhaustively

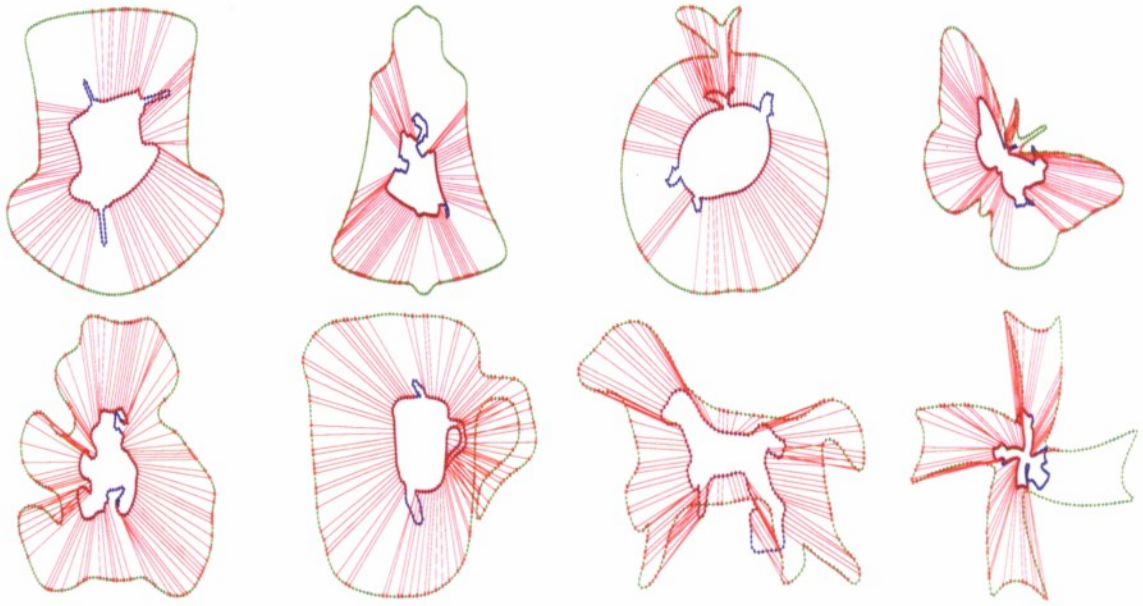

Figure 6: Sample results from the proposed partial shape matching algorithm. One contour (in blue) is shown inside the the other (in green). The matched landmark point pairs are linked by red lines.

checking all possible subwindows, while retaining the global optimality of the result.

In these efficient subwindow search (ESS) algorithms, the searched subwindows are also rectangles, as in the sliding window technique. A rectangular subwindow may not cover the object of interest tightly, which may hurt the object localization accuracy. We adapted the graph-based edge grouping algorithm described in Section 1 to develop a free-shape subwindow search algorithm to address this problem. Besides the preference to cover more positive-score features, we also required the resulting subwindow to align well with edge pixels detected from the image. This way, the boundary of the search subwindow is better aligned with the object boundary and the object localization is more robust against the feature noise. Specifically, we detected a set of disjoint edges in the original image using an edge detector and then formulated the problem of object localization as identifying

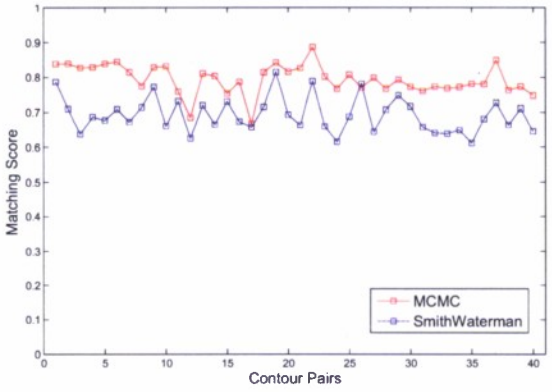

Figure 7: The performance curves of the proposed MCMC-based algorithm and the Smith-Waterman algorithm used in [10].

a subset of edges and connecting them into a closed contour C, so that this contour C minimizes

$$\phi(C) = \frac{|C_G|}{\sum_{f \in C} w(f)}, \tag{4}$$

subject to the constraint

$$\sum_{f \in C} w(f) > 0. \tag{5}$$

where $|C_G|$ is the total length of the gaps along the contour C, and $\sum_{f \in C} w(f)$ is the total score of the features f located inside the contour C. Constraint (5) prevents the detection of an undesired subwindow C that covers all negative-score features and leads to a negative cost $\phi(C)$.

We found that the ratio contour algorithm used in Sections 1 and 2 can be adapted to solve this optimization problem. More specifically, we found that, by removing the constraint (5), the optimal contour minimizing the cost (4) can be found in polynomial time using the ratio-contour algorithm. If this optimal contour satisfies the constraint (5), we proved that this optimal contour is the desired optimal contour C for this image. Otherwise, we can remove the edges in the detected contour and repeat the ratio contour algorithm until finding a contour that satisfies the constraint (5) to obtain an approximate solution. Figure 8 shows several samples results of the developed algorithm and compares it with the result from the ESS algorithm [21].

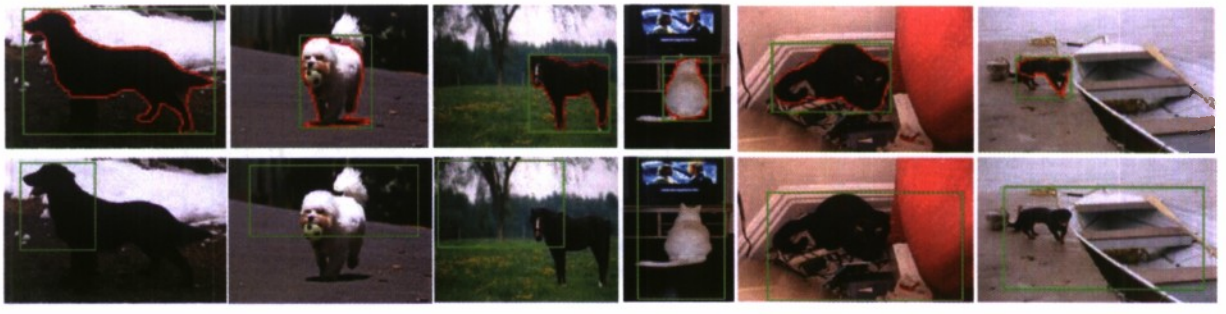

Figure 8: Sample object-localization results of the proposed algorithm (top row) and the ESS algorithm [21] (bottom row). Red contours in the top row are the free-shape subwindows detected by the proposed algorithm and green contours in the top row are the minimum bounding rectangles of the red contours. In this experiment, we use the visual words and feature scores trained in [21].

We tested the proposed algorithm by localizing several categories of animals from the PASCAL VOC 2006 and 2007 databases and comparing its performance with the performance of the ESS algorithm [21]. VOC 2006 database contains 5,304 natural images and VOC 2007 contains 9,963 natural images, where each category of object shows very large variations. Table 2 shows the detection rates of several categories of animals in the VOC 2006 and VOC 2007 datasets using the proposed algorithm and the ESS algorithm.

5 Perceptually Motivated Strategies for Shape Classification

Accurately and reliably measuring the similarity of two shape instances is a fundamental problem in computer vision and plays a central role in many shape-based vision applications including shape matching, shape classification, shape recognition, and shape retrieval. From 2D images, closed contours aligned with object boundaries can be extracted as shape instances, which we also refer to as shape contours. These extracted shape contours may demonstrate a large amount of variation, have

	VOC 2006		VOC 2007	
dataset	Proposed	ESS	Proposed	ESS
dog	0.502	0.458	0.419	0.389
cat	0.524	0.408	0.433	0.422
sheep	0.337	0.281	0.132	0.095
cow	0.436	0.298	0.217	0.176
horse	0.448	0.370	0.398	0.388

Table 2: The performances of the proposed algorithm and the ESS algorithm on VOC 2006 and VOC 2007 datasets.

highly articulated shape parts, involve global and/or local non-rigid deformations, and contain partial occlusions. Even with such complexities, human vision can easily determine whether two shape contours belong to the same shape class. However, developing computational models and methods that can accomplish the same task has proven to be challenging. We developed two perceptually motivated strategies for improving the measure of the shape similarity.

The first strategy aims to better handle the shape contours that contain thin, elongated *strand structures*. Such strand structures may point inward or outward. Two examples of shape contours with outward strand structures are shown in Fig. 9(a) and (b), and an example of a shape contour with inward strand structures is shown in Fig. 9(e).

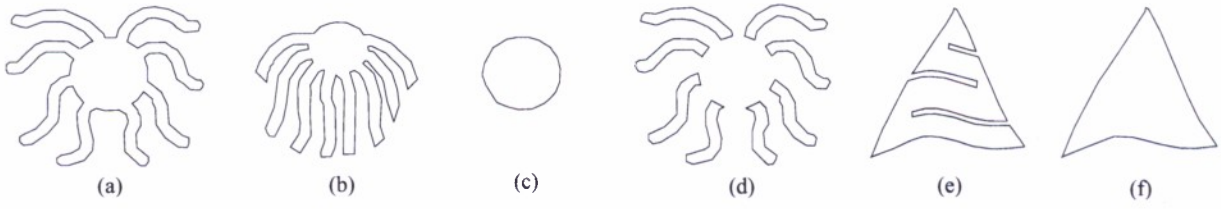

Figure 9: (a-b) Two shape contours with outward strand structures. (c-d) Base structure and strand structures of (a) after shape decomposition. (e) A shape contour with inward strand structures. (f) Base structure of (e) after removing inward strand structures.

In practice, outward strand structures usually describe "leg" or "branch"-like shape components. In human perception, the exact geometry, such as the curvature and length of strand structures, may not be important for shape recognition and classification. For example, the shape contours shown in Fig. 9(a) and (b) are of the same shape class (octopus) and demonstrate high shape similarity in human perception although their legs may be quite different from each other in terms of geometry and size. We developed an algorithm to decompose such a shape contour into a base structure and a set of strand structures, as illustrated in Fig. 9(c) and (d) respectively. When evaluating the similarity between two such shape contours, we match their base structures and strand structures separately. In particular, we apply a deformable shape matching method to compare base structures. When matching strand structures, we simply check whether these two shape contours have a similar number of strands, omitting their detailed geometry.

Inward strand structure can also be extracted by shape decomposition. By removing inward strand structures, we obtain a base structure as illustrated in Fig. 9(f), which is actually the union of the extracted inward structures and the original shape contour. When the inward strand structures are small compared to the structure described by the original shape contour, their removal does not

affect the general human perception of the shape contour. For example, humans usually perceive the shape contours in Fig. 9(e) and (f) to be of the same shape class. We handled such shape contours by extracting and removing the inward structures before shape matching and classification.

The second strategy aims to better handle shape contours that show good bilateral symmetry. For such a shape contour, a certain level of scaling along its symmetric axis or the direction perpendicular to its symmetry axis usually does not change the human perception of its shape. For example, the three different shape contours shown in Fig. 10(a) (b) and (c) all belong to the same shape class (tree) in human perception. We developed an algorithm to identify such symmetric shape contours and unified their aspect ratio before quantitatively evaluating their shape similarity. Here we define the *aspect ratio* of a symmetric shape contour to be the ratio between the length and width of its bounding box along the symmetry axis, as illustrated in Fig. 10(a).

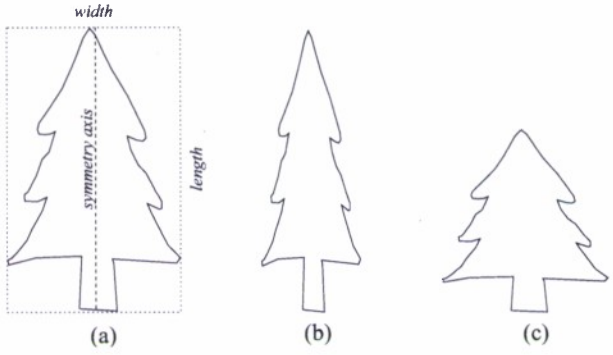

Figure 10: (a) A shape contour with good bilateral symmetry. Its symmetry axis is shown with a dashed line and its bounding box is shown with a dotted line. (b) The shape contour produced by scaling (a) along the direction that is perpendicular to its symmetry axis. (c) The shape contour produced by scaling (a) along its symmetry axis.

To test the proposed strategies, we selected the Inner Distance Shape Context (IDSC) method [24] to measure the shape-matching cost between two base structures. Yang et al [44] developed a new approach to classify a large set of shape contours by extending pairwise shape matching to group-wise shape matching in an unsupervised fashion. For this approach, a locally constrained diffusion process (LCDP) was developed to enhance the similarity of two shape contours if they have low matching cost with another shape contour. This LCDP method also uses the IDSC method for measuring the pairwise shape similarity. LCDP achieves state-of-the-art shape classification performance on several well-known datasets. We also conducted an experiment using the proposed strategies to improve the performance of LCDP, by using IDSC augmented with the proposed strategies as the pairwise shape-matching method.

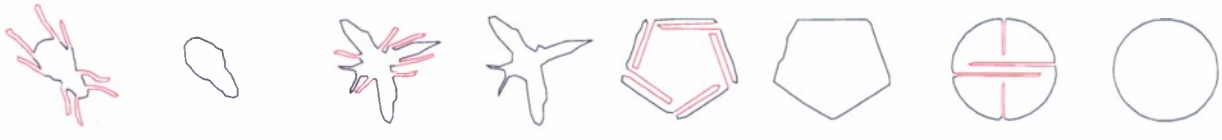

Figure 11: Example strand structures, and base structures found by the proposed method. The *red curves* represent the inward or outward strand structure, and the *black curve* represents the base structure.

Our experiments are based on the widely-used MPEG-7 shape dataset (specifically the MPEG-7 CE-Shape-1 Part B) [22] that defines 70 shape classes, where each shape class contains 20 different

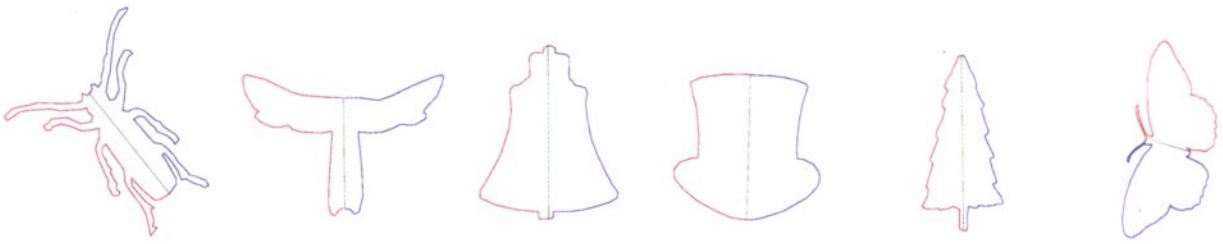

Figure 12: Example symmetric shape contours in the MPEG7 dataset. Symmetric axes are shown in green.

shape contours. We used Bullseye testing to evaluate the performance of the shape classification. In this test, a shape contour is selected from the dataset as the template, and matched to all 1,400 shape contours in this dataset. The 40 most similar shape contours (i.e. with the smallest matching cost) are selected, and out of these 40, we count the number of shape contours that are actually in the same shape class as the template. This number is divided by 20 (the number of shape contours in the template class) to obtain a classification rate. This process is repeated by taking each of the 1,400 shape contours as the template to obtain an average classification rate as the performance. Figure 11 shows several examples of the shape contours in the MPEG-7 dataset that are decomposed into base and strand structures by using Strategy I. Figure 12 shows several examples of the symmetric shape contours in the MPEG-7 dataset as determined by Strategy II. Table 3 shows the Bullseye testing results on the MPEG-7 dataset using the original IDSC method [24], the original LCDP method [44], the IDSC and LCDP methods augmented with the proposed strategies, and other recently published methods. By using the proposed strategies, the shape classification rate of IDSC is improved from 85.40% to 88.39% and the shape classification rate of LCDP is improved from 92.36% to 95.60%.

6 Shape Correspondence and Its Performance Evaluation

Statistical shape modeling provides an effective way to quantitatively describe various shape structures and their possible variations. Accurately identifying corresponded landmarks from a population of shape instances and objectively evaluating the shape correspondence performance are two major challenges in constructing statistical shape models. We developed a new benchmark for shape-correspondence performance evaluation. The system diagram for the proposed benchmark is illustrated in Fig. 13. The benchmark consists of the following five components: (C1) specifying a ground-truth statistical shape model to describe the underlying ground-truth shape space, where we use a point distribution model (PDM) [11] as statistical shape models, (C2) using this ground-truth shape model to randomly generate a set of continuous shape contours $S_1, S_2, \ldots, S_n$, (C3) running a test shape-correspondence algorithm on these shape contours to identify a set of corresponded landmarks, (C4) deriving a statistical shape model from the identified landmark sets, and (C5) assess how well the derived statistical shape model describes the ground-truth shape space defined by the ground-truth statistical shape model. This assessment is achieved by comparing the shape instances sampled from these two shape models using a bipartite matching and other matching methods. This five-step process evaluates a shape-correspondence algorithm's ability to recover the underlying ground-truth shape space in the continuous shape domain. By introducing a ground-truth shape model, the proposed benchmark allows for a more objective evaluation of shape correspondence performance that is landmark independent. The proposed benchmark can easily be extended to 3D cases where each shape instance is a 3D surface.

Method	Rate
Proposed method + IDSC + LCDP	**95.60** %
IDSC + LCDP + unsupervised GP [44]	93.32 %
IDSC + LCDP [44]	92.36 %
IDSC + LP [6]	91.61 %
Contour Flexibility [43]	89.31 %
Proposed method + IDSC	**88.39** %
Shape-tree [16]	87.70 %
Triangle Area [2]	87.23 %
IDSC(EMD) [25]	86.56 %
Hierarchical Procrustes [29]	86.35 %
Symbolic Representation [13]	85.92 %
IDSC [24]	85.40 %
Shape L'Âne Rouge [33]	85.25 %
Multiscale Representation [1]	84.93 %
Polygonal Multiresolution [5]	84.33 %
Fixed Correspondence [37]	84.05 %
Chance Probability Function [36]	82.69 %
Curvature Scale Space [31]	81.12 %
Generative Model [40]	80.03 %

Table 3: Shape classification rate on the MPEG-7 dataset.

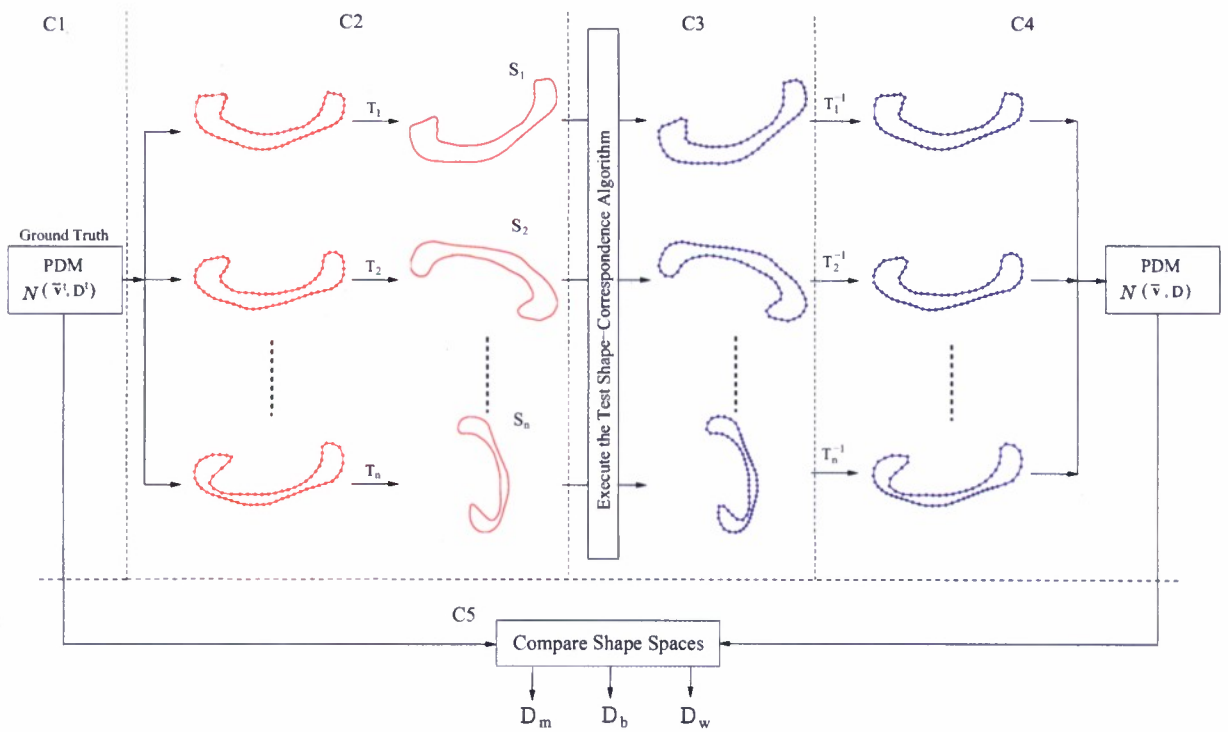

Figure 13: An illustration of the proposed shape-correspondence evaluation benchmark.

In general, shape correspondence methods can be grouped into one of two categories: global methods and pair-wise methods. For global methods [14, 39], an objective function which considers the entire population of shape instances is optimized. For pair-wise methods [8, 34], one shape instance from the population is designated as the template and the remaining target shape instances are optimized to the template one by one. While global methods may produce a more accurate shape correspondence they tend to scale poorly when the population size becomes very large. On the other hand, since a pair-wise method only considers two shape instances at any time, they tend to be less compute intensive and scale favorably to the size of the population. However, because a single template shape instance is chosen from the population, pair-wise methods tend to be less accurate and can perform unsatisfactorily when the population has a large amount of variance.

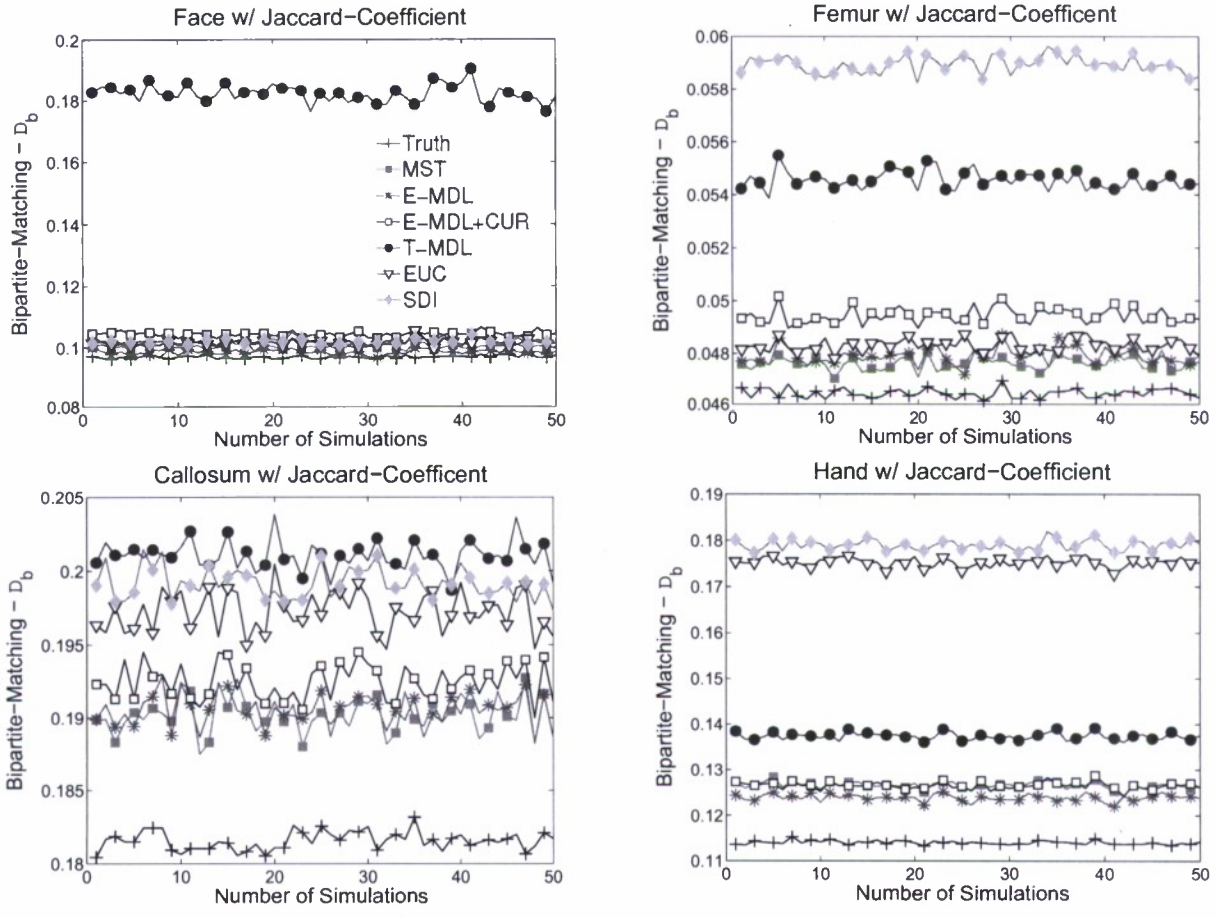

Figure 14: Performance of six shape correspondence algorithms. The x-axis indicates the round of the random simulation. The curves with the "+" symbols are the matching cost between the ground-truth shape model and itself.

To address the limitations of global and pair-wise methods, we developed a new shape correspondence algorithm that pre-organizes the population of shape instances in a tree. Specifically, this is achieved by constructing a minimum spanning tree (MST), where each node represents a shape instance and each edge connects two very similar shape instances. The pre-organization step provided by the MST allows us to incorporate global information about the population of shape instances prior to shape correspondence. A root node is then selected which represents the starting

	MST	T-MDL	E-MDL	E-MDL+CUR	EUC	SDI
Hand	2927	50784	107317	304504	29572	739
Callosum	2318	44732	107506	278832	28420	703
Femur	1757	59663	109875	261093	28538	740
Face	2417	50710	103822	259551	28286	745

Table 4: CPU time (in Seconds) required by the six test shape correspondence algorithms.

shape instance and then, using the constructed MST and the selected root node neighboring shape instances can be corresponded efficiently and accurately using a pair-wise method.

Figure 14 shows the performance of this algorithm (abbreviated as MST) on four datasets using the above mentioned benchmark, with a comparison with other recent shape correspondence algorithms: Thodberg's implementation of the minimum description length method (T-MDL) [39, 38], Ericsson and Karlsson's implementation of the MDL method (E-MDL) [20], Ericsson and Karlsson's implementation of the MDL method with curvature distance minimization (E-MDL+CUR) [20], Ericsson and Karlsson's implementation of the reparameterisation method by minimizing Euclidean distance (EUC) [20], and Richardson and Wang's implementation of a method that combines landmark sliding, insertion, and deletion (SDI) [35]. The performance measure is the bipartite matching cost between two shape spaces and therefore, the lower the better. Table 4 shows the running time of these algorithms on a Linux workstation running Intel Xeon 3.4GHz processor with 4GB of RAM.

References

[1] T. Adamek and N. E. O'Connor. A multiscale representation method for nonrigid shapes with a single closed contour. *IEEE Transactions Circuits and Systems for Video Technology*, 14(5):742–753, 2004.

[2] N. Alajlan, M. S. Kamel, and G. H. Freeman. Geometry-based image retrieval in binary image databases. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(6):1003–1013, 2008.

[3] S. An, P. Peursum, W. Liu, and S. Venkatesh. Efficient algorithms for subwindow search in object detection and localization. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 264–271, 2009.

[4] P. Arbelaez. Boundary extraction in natural images using ultrametric contour maps. In *IEEE Workshop on Perceptual Organization in Computer Vision*, 2006.

[5] E. Attalla and P. Siy. Robust shape similarity retrieval based on contour segmentation polygonal multiresolution and elastic matching. *Pattern Recognition*, 38(12):2229–2241, December 2005.

[6] X. Bai, X. Yang, L. J. Latecki, W. Liu, and Z. Tu. Learning context sensitive shape similarity by graph transduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009.

[7] F. Bookstein. *Morphometric tools for landmark data*. Cambridge University Press, 1991.

[8] F. L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):567–585, 1989.

[9] A. Bosch, A. Zisserman, and X. Munoz. Representing shape with a spatial pyramid kernel. In *Proceedings of the 6th ACM international Conference on image and video retrieval*, pages 401–408, 2007.

[10] L. Chen, R. F. R., and M. Turk. Efficient partial shape matching using smith-waterman algorithm. In *IEEE Computer Vision and Pattern Recognition Workshops*, 2008.

[11] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[12] T. Cour, F. Benezit, and J. Shi. Spectral segmentation with multiscale graph decomposition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.

[13] M. R. Daliri and V. Torre. Robust symbolic representation for shape recognition and retrieval. *Pattern Recognition*, 41(5), 2008.

[14] R. Davies, C. Twining, T. Cootes, J. Waterton, and C. Taylor. A minimum description lengtli approach to statistical shape modeling. *IEEE Transactions on Medical Imaging*, 21(5):525–537, May 2002.

[15] I. L. Dryden and K. V. Mardia. *Statistical Shape Analysis*. Wiley, Chichester, 1998.

[16] P. F. Felzenszwalb and J. D. Schwartz. Hierarchical matching of deformable shapes. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[17] V. Ferrari, L. Fevrier, F. Jurie, and C. Schmid. Groups of adjacent contour segments for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(1):36–51, 2008.

[18] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman and Hall/CRC, Florida, 1996.

[19] K. Grauman and T. Darrell. Pyramid match kernels: Discriminative classification with sets of image features. In *IEEE International Conference on Computer Vision*, 2005.

[20] J. Karlsson and A. Ericsson. A ground truth correspondece measure for benchmarking. In *International Conference on Pattern Recognition*, volume 3, pages 568–573, 2006.

[21] C. Lampert, M. Blaschko, and T. Hofmann. Beyond sliding windows: Object localization by efficient subwindow search. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[22] L. Latecki and R. Lakaemper. Shape similarity measure based on correspondence of visual parts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1–6, October 2000.

[23] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2169–2178, 2006.

[24] H. Ling and D. W. Jacobs. Shape classification using the inner-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(2):286–299, 2007.

[25] H. Ling and K. Okada. An efficient earth mover's distance algorithm for robust histogram comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(5):840–853, 2007.

[26] M. Maire, P. Arbelaez, C. Fowlkes, and J. Malik. Using contours to detect and localize junctions in natural images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[27] D. Martin, C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5):530–549, 2004.

[28] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *IEEE International Conference on Computer Vision*, volume 2, pages 416–423, July 2001.

[29] G. Mcneill and S. Vijayakumar. Hierarchical procrustes matching for shape retrieval. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 885–894, 2006.

[30] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[31] F. Mokhtarian and M. Bober. *Curvature Scale Space Representation: Theory, Applications, and MPEG-7 Standardization.* Kluwer Academic Publishers, Norwell, MA, USA, 2003.

[32] R. Nock and F. Nielsen. Statistical region merging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(11):1452–1458, 2004.

[33] A. M. Peter, A. Rangarajan, and J. Ho. Shape l'ane rouge: Sliding wavelets for indexing and retrieval. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[34] M. Powell. A thin plate spline method for mapping curves into curves in two dimensions. In *Proc. Computational Techniques and Applications*, pages 43–57, 1995.

[35] T. Richardson and S. Wang. Nonrigid shape correspondence using landmark sliding, insertion and deletion. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, pages II–435–442, 2005.

[36] B. Super. Learning chance probability functions for shape retrieval or classification. In *CVRP Workshop*, 2004.

[37] B. Super. Retrieval from shape databases using chance probability functions and fixed correspondence. *International Journal of Pattern Recognition and Artificial Intelligence*, 20(8):1117–1138, 2006.

[38] H. Thodberg. Adding curvature to minimum description length shape models. In *British Machine Vision Conference*, volume 2, pages 251–260, 2003.

[39] H. Thodberg. Minimum description length shape and appearance models. In *Information Processing in Medical Imaging Conference*, pages 51–62, 2003.

[40] Z. Tu and A. L. Yuille. Shape matching and recognition - using generative models and informative features. In *European Conference on Computer Vision*, pages 195–209, 2004.

[41] C. Wallraven, B. Caputo, and A. Graf. Recognition with local features: the kernel recipe. In *IEEE International Conference on Computer Vision*, volume 1, 2003.

[42] S. Wang, T. Kubota, J. Siskind, and J. Wang. Salient closed boundary extraction with ratio contour. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):546–561, 2005.

[43] C. Xu, J. Liu, and X. Tang. 2d shape matching by contour flexibility. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1):180–186, 2009.

[44] X. Yang, S. Koknar Tezel, and L. Latecki. Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 357–364, 2009.

Personnel:

Faculty:

Song Wang*

Graduate Students:

Brent Munsell* (Ph.D. in 2009)
Joachim Stahl* (Ph.D. in 2008),
Habib Moukalled (M.S. in 2009)
Dana Ke (M.S. in 2008)
Andrew Temlyakov* (M.S. in 2007)

Pahal Dalal* (Ph.D. student, started from 2005)
Kenton Oliver* (Ph.D. student, started from 2007)
Andrew Temlyakov* (Ph.D. student, started from 2007)
Yu Cao* (Ph.D. student, started from 2007)
Jarrell Waggoner (Ph.D. student, started from 2007)
Zhiqi Zhang (Ph.D. student, started from 2009)
Habib Moukalled (Ph.D. student, started from 2009)
Dhaval Salvi (Ph.D. student, started from 2009)

* Partially supported by this AFOSR grant.

Publications:

1. Z. Zhang, Y, Cao, J. Waggoner, D. Salvi, S. Wang. Free-Shape Subwindow Search for Object Localization, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, San Francisco, CA, 2010

2. A. Temlyakov, B. Munsell, J. Waggoner, S. Wang. Two Perceptually Motivated Strategies for Shape Classification, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, San Francisco, CA, 2010

3. J. Singletary, P. Wood, J. Du-Quiton, S. Wang, X. Yang, S. Vishnoi, W. Hrushesky. Imaging Multidimensional Therapeutically Relevant Circadian Relationships, *International Journal of Biomedical Imaging*, Article ID 231539, 8 pages, 2009

4. H. Moukalled, D. Deliyski, R. Schwarz, S. Wang. Segmentation of Laryngeal High Speed Videoendoscopy in Temporal Domain using Paired Active Contours, *International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA)*, Firenze, Italy, 2009

5. P. Dalal, L. Ju, M. McLaughlin, X. Zhou, H. Fujita, S. Wang. 3D Open-Surface Shape Correspondence for Statistical Shape Modeling: Identifying Topologically Consistent Landmarks, *IEEE International Conference on Computer Vision (ICCV)*, Kyoto, Japan, 2009

6. B. C. Munsell, A. Temlyakov, S. Wang. Fast Multiple Shape Correspondence by Pre-Organizing Shape Instances, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 840-847, Miami Beach, FL, 2009

7. B. C. Munsell, P. Dalal, S. Wang. Evaluating Shape Correspondence for Statistical Shape Analysis: A Benchmark Study, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(11):2023-2039, 2008 (PDF)

8. J. S. Stahl, K. Oliver, S. Wang. Open Boundary Capable Edge Grouping with Feature Maps, *IEEE Computer Society Workshop on Perceptual Organization in Computer Vision (POCV)*, Anchorage, AK, 2008 (PDF)

9. J. S. Stahl, S. Wang. Globally Optimal Grouping for Symmetric Closed Boundaries by Combining Boundary and Region Information, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(3):395-411, 2008 (PDF)

10. J. S. Stahl, S. Wang. Edge Grouping Combining Boundary and Region Information, *IEEE Transactions on Image Processing*, 16(10):2590-2606, 2007

11. B. C. Munsell, P. Dalal, S. Wang. A New Benchmark for Shape Correspondence Evaluation, *International Conference on Medical Image Computing and Computer Assisted Intervention*, vol. I, pp. 507-514, Brisbane, Australia, 2007

12. P. Dalal, B. Munsell, S. Wang, J. Tang, K. Oliver, H. Ninomiya, X. Zhou, and H. Fujita, A Fast 3D Correspondence Method for Statistical Shape Modeling, *IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, 2007

13. F. Ge, S. Wang, and T. Liu, New Benchmark for Image Segmentation Evaluation, *Journal of Electronic Imaging*, 16(3):033011, 16 pages, 2007

14. S. Wang, J. S. Stahl, A. Bailey, M. Dropps. Global Detection of Salient Convex Boundaries, *International Journal of Computer Vision*, 71(3):337-359, 2007

Dissertations:

Two Ph.D. dissertations are produced under the support of this AFOSR grant. These dissertations are publicly available at the University of South Carolina Library.

1. Combining Boundary and Region Information for Perceptual Organization, Joachim Stahl, Ph.D., 2008

   Abstract: Perceptual organization, or grouping, is an important problem in computer vision and image processing that seeks to identify perceptually salient structures in noisy images. As an important step in mid-level computer vision, grouping can provide useful input to many high-level computer-vision applications such as object recognition or content-based image retrieval. To identify a salient structure, a set of tokens is first obtained from the original image, and then a subset of these tokens is grouped that minimizes a cost function (or maximizes saliency). This work introduces a series of new edge grouping methods to detect perceptually salient structures in noisy images, where the grouping tokens are edge segments. Unlike previous edge grouping methods, which base their saliency measures exclusively on boundary properties, the proposed methods incorporate region information into their saliency measure. The use of region information makes these methods more robust to noise in the image, and add capabilities such as targeting structures with specific region characteristics. The first method presented introduces the general problem of incorporating region information into an edge grouping method. The second method targets structures that are a priori known to be convex. The third method uses

symmetric trapezoids as its grouping tokens to target structures that are a priori known to have good bilateral symmetry. The fourth method extends the first method with the capability of detecting open boundaries (for structures not completely present within the perimeter of the input image). To find the optimal grouping with the minimum cost, a special graph model is developed in each case and the grouping problems are reduced to finding a special kind of cycle in these graphs. This optimal cycle-finding problem can be solved in polynomial time using a known graph algorithm. The presented methods are tested on both synthetic data and real images, and their performance is compared against previous edge-grouping methods. Some major results of this dissertation are summarized in Sections 1 and 2 of this report.

2. Shape Correspondence for Statistical Shape Modeling: Algorithms and Performance Evaluation, Brent Munsell, Ph.D., 2009

   Abstract: In order to accurately measure structural shape and its possible variation, statistical shape analysis has become a major research topic in computer vision and medical image analysis in recent years. In statistical shape analysis a population of shape instances is given where each shape instance is in the form of a smooth 2D contour or a smooth 3D surface. The goal is to construct a statistical shape model that accurately captures the variability of the given shape structure described by the population of shape instances. In constructing a statistical shape model the first step is to identify a set of landmarks for each shape instance, where a landmark is defined as a point of correspondence across the population that can be used to examine and measure shape change. In general these landmarks can be identified manually by a human (expert), or automatically via software. Manually identifying corresponded landmarks can be achieved, however such a method is both subjective and error prone. Because of this, developing more accurate and efficient shape correspondence methods that automate the landmark identification process has been widely investigated over the last several years. Even though much progress has been made, the development of an efficient and accurate shape correspondence method that scales favorably to the size of the population is still a largely unsolved problem. Another open problem in statistical shape analysis is the objective evaluation of these shape correspondence methods. One major reason is the unavailability of a ground-truth shape correspondence, which would be defined by a group of experts that manually identify the corresponded landmarks. Currently, this limitation is addressed by three general measures that are used to evaluate the shape correspondence performance. These three measures describe the properties of the statistical shape model constructed from a shape correspondence result and not against some known ground-truth. The research presented in this dissertation attempts to address these two problems by developing an efficient and accurate shape correspondence method that scales well to the size of the population, and develop a shape correspondence benchmark to objectively evaluate shape correspondence performance against some known ground-truth. Major results of this dissertation are summarized in Section 6 of this report.